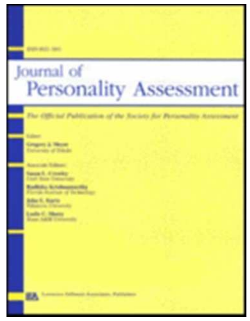
A Comparison Between Comprehensive System and an Early Version of the Rorschach Performance Assessment System Administration with Outpatient Children and Adolescents.

Published version:

DOI:10.1080/00223891.2014.889700

(Article begins on next page)

Reese J; Viglione DJ; Giromini L. A Comparison Between Comprehensive System and an Early Version of the Rorschach Performance Assessment System Administration with Outpatient Children and Adolescents..
JOURNAL OF PERSONALITY ASSESSMENT. 96 (5) pp: 515-522.
DOI: 10.1080/00223891.2014.889700



A Comparison Between Comprehensive System & Rorschach Performance Assessment System Administration with Outpatient Children and Adolescents

Journal:	*Journal of Personality Assessment*
Manuscript ID:	JPA-2013-296.R1
Manuscript Type:	General Submission
Keywords:	Rorschach < Measures, Child, Responses, R-PAS, Assessment

Abstract

For many years, the effects of variability in the length of Rorschach records has been debated, and a new administration procedure aimed at reducing the proportion of short and long records has recently been introduced. Using an outpatient sample of children and adolescents, the current study explored the impact of this new administration, found in the Rorschach Performance Assessment System (R-PAS), on the central tendencies of Rorschach variables. Specifically, the mean values of 51 variables in 142 Comprehensive System vs. 99 R-PAS collected records were compared with each other. Results showed comparable mean values across CS and R-PAS administration methods for the variables that guide interpretation with children and adolescents. Both methods produced an equivalent number of long (>27 responses) records. The only relevant difference that emerged is that the R-PAS administration yielded significantly fewer short (14-16 responses) records and lower variability in the number of responses.

Keywords: Rorschach; Child; Responses, R-PAS; Assessment

A Comparison Between Comprehensive System & Rorschach Performance Assessment System Administration with Outpatient Children and Adolescents

For many years, the effects of variability in the length of Rorschach records, as defined by the number of responses (R) has been debated (e.g., Cronbach, 1949; Fiske & Baughman, 1953). Some have recommended that R be controlled to improve the psychometric efficiency and usefulness of the test (Cronbach, 1949; Holtzman, 1958; Meyer, 1992). When Exner created the Comprehensive System (CS) in 1974, he considered but decided not to restrain R because his data revealed less variation in R than others had found. As a result, like most Rorschach systems, the CS provides few tactics to constrain R. Later, because research revealed that test-retest reliability of records with less than 14 responses are limited (Exner, 1988), a minimum of 14 responses was required along with a re-administration procedure for cases when less than 14 responses were procured. Since that time, in almost all cases efforts to manage or increase the number of responses are limited to prompting for a second response on Card I or re-administering the test if fewer than 14 responses are offered, because of the current fact that 14 responses is required in the CS (Exner, 2003).

Nevertheless, recent research and reviews (Meyer, 1992; Meyer, Erdberg, & Shaffer, 2007; Shaffer, Erdberg, & Haroian, 2007; Viglione & Hilsenroth, 2001; Viglione & Meyer, 2008) have re-opened the question of constraining R. First of all, the variation in R in most samples is much greater than Exner's initial data would have suggested. For instance, across nonpatient samples from 17 countries the average standard deviation for R was 7.9 (Meyer et al., 2007), whereas the CS normative data (Exner, 2003) standard deviation is only 4.4. If one were to assume a normal distribution for R among the Meyer et al.'s international data, its mean of

22.3 would suggest that 19% of the records would have 14 or 15 responses and 18% would have 30 responses or more. Assuming the same for the CS normative sample (M = 22.3) would instead lead to only 5% and 6% respectively. Thus far more short and long records than initially expected are typically produced.

In addition, some authors have claimed that variations in R may pose problems for the interpretation of the large number of Rorschach variables that are correlated with R (Viglione & Meyer, 2008). For example, in a large (N= 1,342) mixed sample of patients, non-patients and offenders, for records with 14 to 17 responses, the average for uncommon detail (Dd) was 1.6 and for Experienced Stimulation (es) 6.1, whereas the means for these same variables was 7.1 and 14.5 in records with 28 or more responses (Viglione & Meyer, 2008). With such dramatic swings, one may not be able to determine whether any particular low or high score is associated with the underlying construct or the number of responses, thus potentially hampering interpretive accuracy and certainty. Supportive of these inferences about short records, Dean, Viglione, Perry, and Meyer (2007) demonstrated that prompting for more responses in a clinical setting prone to frequent low R protocols maintains or improves the validity of Rorschach measures of thought disorder and psychosis while reducing the proportion of brief records. Such a result is also in line with Meyer's (1993) empirical findings obtained within an inpatient, adult sample (N = 90), which indicated that the Schizophrenia Index (SCZI; Exner, 1986, 1991) "is most able to differentiate patients with a psychotic disorder when protocols are of average length. If protocols have many or few responses, the SCZl is no longer effective because the formal cause nature of scoring principles results in frequency criteria that are less valid representations of their underlying constructs" (p. 170).

This compilation of findings led Meyer, Viglione, Mihura, Erdberg and Erard (2011) to recommend small changes in Rorschach administration for the Rorschach Performance Assessment System, their recently published new Rorschach method. R-PAS administration includes (1) encouraging examinees to give 2 or 3 responses per card, (2) prompting for at least two responses to each card, and (3) preventing more than four responses being given to any card. Thus, R-PAS does not permit records with more than 40 responses over the 10 cards.

Some Rorschach authorities contend that modifying the administration procedures of the Rorschach would have negative effects on the test. Mattlar (2011), for example, stressed that the R-PAS administration violates one of the leading principles in the CS, specifically that the examination is non-directive, and the examiner interferes as little as possible with the examinee's work. Similarly, Weiner (2012) recently claimed that changing the administration method of the Rorschach to the R-PAS administration implies a radical change in the test, from a free-choice method to a forced-choice method. This point of view is also in line, to some extent, with some empirical findings demonstrating the clinical utility of R. Lipgar and Waehler (1991), and Perry and Kinder (1990), for example, showed that the total number of R, as well as the number of responses per card, differentiate examinees in interpretively important ways. Perhaps more importantly, Exner himself believed that R has an interpretative importance, and in fact he included this variable for the prediction of suicide (Exner, 1986).

Independent of these varying opinions and claims regarding number of responses and administration procedures remains the fact that very few studies have investigated the impact of R-PAS administration. Further, none of these studies have used experienced examiners in applied settings with clinical patients. Therefore, further investigation of this new administration procedure and its resulting impact on the Rorschach is sorely needed. One might wonder, for example, to what extent will this new procedure affect the production of Rorschach responses? Will the R-PAS administration lead to more simplistic responses, as a result of the task demands for more responses?

These questions may be of particular importance especially when testing children and adolescents. Indeed, there is some evidence that children and young adolescents tend to produce more simplistic responses relative to adults. According to several CS normative studies published over the past 40 years (Exner, 2003; Exner, Thomas, & Mason, 1985; Exner & Weiner, 1982, 1994; Stanfill, Viglione, & Resende, 2013; Weiner, 2003), younger individuals indeed tend to produce more A (animal content), C and CF (color dominated responses), and DQv (vaguely outlined responses), and less M (human movement), FC (form dominated color responses), and DQ+ (synthesis responses). In other words, children and adolescents tend to produce less complex and less sophisticated responses when compared to those of adults. Even before these CS studies, Ames and colleagues (Ames, Metraux, Rodell, & Walker, 1974; Ames, Metraux, & Walker, 1971) had already reported that most of the differences among children and adolescents at different ages occurred for variables such as R, D% (percentage of common easily seen detail response), Pure F% (percentage of pure form responses), M, W (whole locations),and FC (form dominated color responses): CF (color dominated responses), thus indicating that the younger the respondent, the less complex and sophisticated the responses. Accordingly, investigating the extent to which R-PAS leads to more simplistic (and thus less clinically useful) responses, as a result of the directive to give more responses, is particularly relevant for children and adolescents, as they already tend to produce fewer clinically useful responses than adults.

To add to the literature on these under-investigated topics, in this archival, exploratory study we applied a non-randomized, quasi-experimental design, and compared the Rorschach responses produced by two clinical groups of children and adolescents: one group was administered the Rorschach test according to the standard, CS procedure (control group); the other group was administered the Rorschach test according to the new R-PAS administration (experimental). Means and standard deviations of all Rorschach variables included in the R-PAS were compared, and several other analyses were undertaken.

Hypotheses

To date, very few studies have empirically investigated the effects of the R-PAS administration on the Rorschach scores, and none – to the best of our knowledge – used clinical children or adolescents. This study was therefore mainly exploratory. Based on the scarce, available literature, however, we did have some tentative hypotheses. First, given the more explicit instruction about the number of responses required by the examinee, we predicted a significant reduction in the number of short and long records, and lowered variability in the total number of responses (R) when comparing R-PAS to CS protocols. Importantly, in line with Meyer et al. (2011), R-PAS administration was not expected to impact the central tendency of R, but only its variability.

Furthermore, we anticipated that the effect of instructions and prompts with R-PAS administration would be shared by all variables so that very few, if any, variables would demonstrate significant mean differences. Indeed, despite the R-PAS administration being expected to lower the variability of R (i.e., to reduce the number of short and long records), there was no statistical reason to believe that such an effect would impact the central tendency of any Rorschach variables, as differences in variability do not imply, per se, any changes in central tendency. Thus, for all Rorschach variables under consideration we anticipated potential differences in the variability values (possibly more marked for those variables correlated with R), but did not predict any significant differences in the mean values.

Based on previous work, theoretical considerations, and concerns about R-PAS administration being more directive than optimal, we also thought it was possible that task demands for more responses might lead to more easy to give, simple responses with the R-PAS administration, so that it might produce more common detail (D), simplified form responses (F), animal content responses (A), and popular responses (P), and fewer synthesized responses (Sy), and responses with two or more determinants (Blend). This might result in simpler responses overall (Complexity), especially for 2nd and 3rd responses to cards, which are those provoked by the method. Also, given that there is more consistency across cards with R-PAS administration, we expected the number of responses to the last three colored cards (R8910% in R-PAS, Afr in the CS) to be lower. These last three cards produce more responses than do the other cards with CS administration. Possibly, this modified administration equalizes the number of responses across cards with the effect of lowering Afr.

For the rest of the responses, we expected no differences between the two groups. Accordingly, we anticipated that the number of significant differences would not exceed chance and any such differences would likely be small, with Cohen's *d*'s of about .4 or less, with the possible exception of Afr.

Method

Participants

A sample of 241 de-identified, archival Rorschach protocols from outpatient children and adolescents was extracted from computerized records. All were outpatient clinical evaluations that included the Rorschach completed between March 2007 and June 2011 at an outpatient community mental health agency based in Canton, Ohio. The agency serves approximately 2,500 individuals each year, 86% of which have Medicaid and lower socioeconomic means. Approximately 25% of the agency's clients are referred for psychological evaluation with the goal of diagnostic clarification. A subset of this group participates in a Rorschach administration, which occurs toward the end of the overall assessment battery in order to ensure maximum rapport has been established.

There were 152 boys and 89 girls, aged 5 to 16 years (M = 12.3, SD = 3.0) at the time of administration. With regard to racial identity, 77.2% Caucasian, 13.7% African American, .8% Hispanic, .4% Asian American, 7.9% Other (Multiracial). All were English speaking. The individuals were organized within the database according to the primary diagnosis generated by the psychological evaluation. Approximately 18.7% had a primary diagnosis of Attention Deficit Hyperactivity Disorder (ADHD), 11.6% had an anxiety disorder (excluding Post-Traumatic Stress Disorder; PTSD), 4.6% had an Autism Spectrum Disorder, 7.1% had a primary behavior disorder diagnosis (excluding ADHD), 25.7% had a mood disorder, .8% had a psychotic disorder, 14.9% had PTSD, and 16.6% were court-involved due to sexually inappropriate behavior.

Rorschach and Contrast Groups

Rorschachs were administered and scored by individuals holding graduate degrees in psychology including pre-doctoral psychology interns. All were trained in CS techniques. A single staff psychologist (Reese) conducted the training and supervision of these individuals during the entire sampling period. Any protocols not administered by this supervisor were scored independently by both the examiner and the supervisor, with any differences reconciled through reconsideration and discussion. Twenty-four different examiners contributed data during the sampling period.

Contrast groups for our study are based on the method of administration and date of administration. All protocols administered before April 1, 2010 (N = 142) used standard CS administration, whereas those collected afterward (N = 99) used R-PAS administration. Importantly, this change in the administration procedure was a natural transition, as the agency implemented the R-PAS administration without any plan for involvement in research. This ensures that all Rorschach coders were blind to the hypotheses of our study. Also importantly, despite the differences in the administration procedures, all records of both the groups were coded according to the CS method (and thus using the same, CS FQ tables), so that in this study the scoring method is not a confound.

There were no significant differences between the two groups in terms of age, gender, years of education, SES, or race (table 1). Instead, a significant difference in terms of diagnostic category, $\chi^2 = 17.7$, $p = .01$, was observed. Examination of the standardized residuals, however, indicated that the two groups only differed as for the presence of court-involved sexual behavior problem children, which was higher within the R-PAS sample when compared to the CS sample. When excluding this subgroup from the analysis, in fact, no other diagnostic differences approached statistical significance. To take into account the possible impact of this difference on the results, all statistical analyses were initially performed with the entire sample, and then repeated after removing the data of all sex offenders. Since no changes in findings occurred, we only discuss findings obtained when considering the entire sample.

We chose to examine variables which have demonstrated sufficient validity, so that they have been adopted for this and other reasons by R-PAS (Meyer et al., 2011). In R-PAS output the variables are differentiated into Page 1 and Page 2 variables, with Page 1 variables having more psychometric support and being more important in interpretation. This selection is strongly influenced by a recent, thorough meta-analysis (Mihura, Meyer, Dumitrascu, & Bombel, 2013). Although some of these variables have only been introduced by the R-PAS and were not coded in the CS (e.g., Aggressive Content, see Gacono & Meloy, 1994), most of them in fact overlap with, or can be easily calculated just by using CS codes. For example, the R-PAS variable MC (Meyer et al., 2011) is identical with the CS variable Experience Actual (EA; Exner, 2003), and the R-PAS code Synthesis (Meyer et al., 2011) corresponds to the CS code of Developmental Quality "+" or "v/+" (DQ+ or DQv/+ respectively, in the CS; Exner, 2003). Because all records included in this study were scored according to the CS guidelines (see above), we could not analyze the "new" variables introduced by the R-PAS and not used in the CS. Thus we only examined those R-PAS Page 1 and Page 2 variables that either overlap, or can be easily calculated just by using CS codes (for additional information about R-PAS terms and their CS counterparts, see Meyer et al., 2011, Appendix D).

Defining cutoffs for "short" and "long" is to some degree arbitrary, so we considered both CS and R-PAS guidelines. According to the Exner text for the CS (2003), "protocols of 14, 15, or 16 answers are usually difficult for interpreters to glean a full picture of the psychology of the client, and often are marked by some situationally related resistance" (p. 53). Along the same lines, Weiner (2003) recommended that examiners "note initially whether the respondent has produced a "short" record", and specifies that "short records consist of 14 to 16 responses, which makes them long enough to be valid but not necessarily full enough to be revealing" (p. 67). Consistent with these assertions, we considered records with $R < 17$ as undesirably "short." Although technically valid, indeed, Rorschach experts consider these records may be less revealing and possibly more misleading than longer records. As for the opposite side of the distribution of R, in this study we considered records with R > 27 as "long", based on Meyer et al.'s (2011) definition of middle range of R as being approximately from 18 to 27. As opposed to those in the middle range of R, records with R > 27 increase labor and time, undermining that the benefits of administering a Rorschach outweigh the costs of it.

To establish inter-rater reliability, we selected 20 records at random which were scored by two raters blind to the other's coding. Ten of the records had previously been independently coded by both the examiner and the supervisor (Reese), and ten more were re-coded by an advanced graduate student, who has coded more than 100 Rorschach records. For these records, the two-way random effects model single measures intraclass correlation coefficients (ICCs) ranged from .60 to 1.00. The mean ICC was .90, the median .92, and according to the suggested benchmarks (Cicchetti, 1994; Shrout & Fliess, 1979), all Rorschach variables included in the analysis demonstrated either excellent (i.e., ICC ≥ .75; 44 variables) or good (i.e., ICC between .60 and .74; 7 variables) inter-rater reliability. Detailed inter-rater reliability information for all variables under investigation is reported in table 2.

Results

Distribution of R

Both administration methods yielded a minimum of 14 responses for all records. Looking at the distributions more closely, however, the R-PAS administration produced only 1 record (i.e., 1% of the total R-PAS sample) with fewer than 17 responses, whereas 22 such short records (i.e., 15.5% of the total CS sample) were obtained with the CS. The R-PAS administration, therefore, yielded significantly fewer short records, Phi = .24, $p < .001$. On the other hand, both methods produced a comparable proportion of long (R>27) records.

The mean number of responses per protocol was 24.3 (SD = 8.4) for the CS records (N = 142) and 24.7 (SD = 5.4) for the R-PAS ones (N = 99). No significant differences in the means were found, t(237.7) =.5, p = .618 (for details see Table 3). According to Levene's test and as expected, the variance of R of the CS records was greater than the R-PAS records, F(1,239) = 14.2, p < .001.

Impact of Administration Method on Page 1 and Page 2 R-PAS Variables

The mean values of Page 1 and Page 2 R-PAS variables are presented in Tables 4 and 5. Almost all variables show strikingly similar mean values for CS and the R-PAS administrations. The average of the absolute value of the 51 effect sizes is a very small *d* = .11 (1 T score points, or 1.5 standard score points) with a standard deviation of .08.

No comparisons are statistically significant when considering uncorrected p-values of .01, nor when adopting Bonferroni's correction for a p-value of .05. By adopting an uncorrected alpha value of .05, only two comparisons are significant differences, with (CF+C)/SumC and PER showing higher mean values within the CS as compared to the R-PAS records. Such a finding of only two positive results among 51 is likely due to chance. In addition, the effect sizes of these results are quite small, with Cohen's *d* being .34 for (CF+C)/SumC, and .26 for PER. Said differently, these differences are in the range of only about 2 or 3 T score points or 3 or 4 standard score points for the normative transformations with R-PAS.

Bayesian Analyses. Although Bayesian analyses are still rare in psychological literature, there is a growing consensus among statisticians that classic null-hypothesis significance tests (NHSTs) underestimates support for the null hypotheses, and overstate the evidence against them (e.g., Berger & Sellke, 1987; Edwards, Lindman, & Savage, 1963; Goodman, 1999; Rouder & Morey, 2011; Rouder, Speckman, Sun, Morey, & Iverson, 2009; Sellke, Bayarri, & Berger, 2001; Wagenmakers, 2007; Wagenmakers & Grünwald, 2006). Put simply, NHSTs seem to work fine when the null hypothesis is false, but encounter serious problems when the null hypothesis is true. Indeed, if the null hypothesis is false, as the sample sizes increase the p-values decrease (as one should expect), but if the null hypothesis is true, increasing the sample sizes does not affect the p-values (see for example, Rouder et al., 2009). That is, with NHSTs increasing the sample size does not allow a researcher to gain evidence for the null hypothesis. In addition, as shown in Rouder and Morey (2011), for a sample size of 500 and an effect size of .2, very small p-values are the norm, i.e., p-values between .04 and .05 are about 10 times more likely under the null hypothesis than under the alternative hypothesis. Notwithstanding, the null hypothesis would still be rejected, according to NHST. This paradoxical behavior of NHST, for which the null hypothesis is rejected also when evidence clearly favors it, is known as the *Lindley's paradox* (Lindley, 1957).

Following this line of reasoning, we implemented the Bayesian procedures as more accurate test of the null hypothesis that R-PAS and CS administrations would produce similar means. For each comparison, we calculated the ratio of the probability of obtaining our data under the null hypothesis to the probability of obtaining our data under the alternative hypothesis, i.e., Pr (data | H_0) / Pr (data | H_1). This ratio is often denoted by *B* and termed the *Bayes factor* (Jeffreys, 1961; Kass & Raftery, 1995). According to Jeffreys (1961), if *B* is greater than 3 (i.e., the null is 3 times more probable of the alternative, given the data) there is "some evidence" for the null, if it is greater than 10 there is "strong evidence" for the null, and if it is greater than 30 there is "very strong evidence" for the null. Vice versa, if *B* is smaller than .33 (i.e., the alternative is 3 times more probable of the alternative, given the data) there is "some evidence" for the alternative, if it is smaller than .10 there is "strong evidence" for the alternative, and if it is smaller than .03 there is "very strong evidence" for the alternative. To compute the *B* values, we used procedures described by Rouder et al. (2009) and used the web-based program provided by the authors. In this approach, the *B* values are calculated according to Rouder et al.'s (2009) equation 1 for the two-sample case, and are termed *JZS B*. As compared to other methods of calculating *B,* the *JZS B* has several advantages: "It makes intuitive sense, it has beneficial theoretical properties, it is not dependent on the measurement scale of the dependent variable, and it can be conveniently computed" (Rouder & Morey, 2011, p. 685).

When calculating the *JZS B*, Rouder et al. (2009) recommend to set $r = 0.5$ as a default, when small differences are of theoretical importance. In this approach, *r* is a scale factor which affects the *JZS B* formula so that the higher its value the more the *JZS B* will tend to provide support for the null (for details, see Rouder et al. 2009). To be appropriately conservative in respect to our hypotheses of equivalence between the R-PAS vs. CS means and to recognize that small differences in norms might make a difference in interpretation, we set *r* to 0.5.

As shown in the last column of Tables 4 and 5, for almost all comparisons the *JZS B* values were greater than 1, thus indicating that the null hypothesis is more probable than the alternative, given the data. Specifically, for 48 variables the *JZS B* value is greater than 1, for one variable it is equal to 1 (thus indicating equal probability for null and alternative, given the data), and for 2 variables (i.e., (CF+C)/SumC and PER) it is lower than 1 (thus indicating that the alternative is more probable than the null, given the data). According to the aforementioned Jeffreys criteria (Jeffreys, 1961), for 31 variables (17 on Page 1, and 14 on Page 2) there is "some evidence" for the null hypothesis, and none of the 51 variables under examination reached the cut-off of *JZS B* $\leq .33$, required to provide "some evidence" for the alternative.

Administration Method and Complexity

Despite our concerns, neither the complexity of the responses (see, for example, Complexity, F%, and Blend in Table 4) nor the proportion of responses to the last three colored cards were significantly different between the two samples. In addition, inspection of *JZS B* values reveals that the null hypothesis is about 5 times more probable than the alternative hypotheses, given the data.

To further investigate whether the R-PAS administration might lead to simple, easy to give, responses, we computed additional analyses which excluded the 1st responses to each card. As previously stated, indeed, we anticipated that the R-PAS administration would possibly lead to an overall lower complexity, i.e., more simple responses, fewer synthesized responses or blends, and to more common D location, pure Form, popular, and A content responses on the 2nd and 3rd responses to cards, as these are the ones that are provoked by the method. Table 6 shows that no significant differences were observed when either the 1st responses to each card were excluded or were they included in the analysis. In addition, *JZS B* values indicate that the null is about 2 to 5 times more probable than the alternative, given the data, for all comparisons. In fact, the complexity and the number of D, F, Sy, Blend, Popular, and animal content responses were very similar across the two samples in both the conditions.

Discussion

To investigate the impact of the R-PAS administration on the Rorschach evaluation of outpatient children and adolescents, we compared the mean values of 51 Rorschach variables produced by 142 CS collected vs. 99 R-PAS collected records. With a p-value of .01, none of the numerous comparisons resulted in statistically significant mean differences. Of note, both the CS and R-PAS administrations produced valid records ($R \geq 14$) in all cases, and a similar number of long records, although – as expected – the R-PAS administration did yield significantly fewer short records and lower variability in the number of responses. As a whole, the main thrust of this study is that CS and R-PAS administration methods yield similar mean values for children and adolescents for the variables that guide interpretation. Thus, on one hand, the R-PAS administration did not lead to more simplistic responses. On the other hand, by no means from this research can one conclude that one administration method has an advantage over the other. This questions awaits for the research.

According to our results, the R-PAS administration has the potential to decrease the likelihood that short records will be produced. In fact, although neither administration methods produced records with less than 14 responses, only 1% of the R-PAS records resulted in short protocols (i.e., with R < 17), compared to 15.5% of the CS records. Given the similar means for variables across the two methods, however, this is unlikely to result in interpretive differences. As reviewed in the introduction section, some authors might interpret the reduced number of short records as a potential strength in terms of validity or utility of the R-PAS method, while others might consider it as a potential limitation. Future research should be undertaken to determine which of these two positions is true, from an objective, empirical point of view. Regardless of the possible interpretations, however, this study is the first to investigate the potential impact of the attempts to constrain the variability of the number of responses within R-PAS administration on a clinical sample of children and adolescents.

In addition to the classic NHST, to investigate the hypothesis that the R-PAS administration does not affect the mean values of the Rorschach variables when compared to the CS administration, this study also implemented a Bayesian approach. Bayesian statistics are not used very often in the Rorschach literature, perhaps because the theory and calculations behind them may appear complex at first sight. However, as discussed previously, the *JZS B* are conveniently computed by using a web-based program, and there is a growing consensus among statisticians that the NHSTs encounter serious problems when testing the null hypothesis. In addition, finding nonsignificant results when comparing two means (i.e., finding p-values above .05) does not provide support to the null hypothesis, but only adds evidence that the null hypothesis cannot be rejected (Altman & Bland, 1995). Thus, aiming to test the (null) hypothesis that the means produced by R-PAS vs. CS administrations do not differ, adopting a Bayesian approach ensured a more appropriate understanding of the data than the more commonly utilized NHST.

This study also provides some information concerning inter-rater reliability. As the R-PAS is new, prior to this research only one study had provided information on the inter-rater reliability of this system: by selecting 50 Rorschach records from ongoing research projects using R-PAS administration, Viglione, Blume-Marcovici, Miller, Giromini and Meyer (2012) found that the great majority of codes obtained good to excellent inter-rater reliability. The current study, from this point of view, confirms and extends such findings, providing additional information in support of the strong inter-rater reliability of the test.

Strengths of this study include the use of an outpatient clinical population undergoing psychological assessments and a natural transition to the R-PAS administration procedure, so at to increase ecological validity. In other words, the agency implemented R-PAS after it was formally introduced at a conference in March 2010 without any plan for involvement in research. The number of examiners contributing data to this study and the good (7 variables) to excellent (44 variables) inter-rater reliability are also strengths. As 24 different examiners administered the Rorschachs used in our sample, there is good reason to believe that these results will generalize across other examiners as well. Lastly, the sample size (*N* = 241) was a strength of this study as a group of this size led to good statistical power and sensitivity to small differences.

Nevertheless, some limitations deserve mentioning and warrant caution when interpreting the results. First, our study used a research design (non-randomized quasi experimental) that did not allow for randomization of the sample over CS vs. R-PAS groups. Thus, it is possible that our results have been affected, to some extent, by unrecognized and uncontrolled confounds. Second, almost all of the examiners were women (23 of the 24) and all of the participants were drawn from one agency—characteristics which may limit the generalizability of the findings. Further, the participants were primarily lower SES, therefore children and adolescents from other SES categories were not well-represented.

References

Ames, L. B., Metraux, R. W., Rodell, J. L., & Walker, R. B. (1974). *Child Rorschach responses: Developmental trends from two to ten years*. New York: Brunner/Mazel.

Ames, L. B., Metraux, R. W., & Walker, R. B. (1971). *Adolescent Rorschach responses: Developmental trends from ten to sixteen years*. New York: Brunner/Mazel.

Altman, D. G., & Bland, J. M. (1995). Absence of evidence is not evidence of absence. *British Medical Journal, 311*, p.485.

Aronow, E., Reznikoff, M., & Moreland, K. (1994). *The Rorschach technique: Perceptual basics, content interpretation, and applications*. Needham Heights, MA: Allyn & Bacon.

Berger, J. O., & Sellke, T. (1987). Testing a point null hypothesis: The irreconcilability of p values and evidence. *Journal of the American Statistical Association, 82*, 112–122.

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment, 6*, 284-290.

Cronbach, L. J. (1949). Statistical methods applied to Rorschach scores: A review. *Psychological Bulletin, 46*, 393-429.

Dean, K., Viglione D.J., Perry, W., & Meyer G.J. (2007). A method to optimize response range while maintaining Rorschach Comprehensive System validity. *Journal of Personality Assessment, 89(2)*, 149-161.

Edwards, W., Lindman, H., & Savage, L. J. (1963). Bayesian statistical inference for psychological research. *Psychological Review, 70*, 193–242.

Exner, J. E. (1974). *The Rorschach: A comprehensive system. Vol. 1: Basic foundations*. New York: Wiley.

Exner, J. E., Jr. (1986). *The Rorschach: A comprehensive system (Vol. 1, 2nd ed.)*. New York: Wiley.

Exner, J. E., Jr. (1988). Problems with brief Rorschach protocols. *Journal of Personality Assessment, 52*(4), 640-647.

Exner, J. E., Jr. (1991). *The Rorschach: A comprehensive system (Vol. 2, 2nd ed.)*. New York: Wiley.

Exner, J. E. Jr. (2003). *The Rorschach: A comprehensive system. Volume 1: Basic foundations (4th ed. Vol. 1)*. New York: John Wiley & Sons.

Exner, J. E., Jr., Thomas, E. A., & Mason, B. (1985). Children's Rorschachs: Description and prediction. *Journal of Personality Assessment, 49(1)*, 13-20.

Exner, J. E., Jr., & Weiner, I. B. (1982). *The Rorschach: A comprehensive system, volume 2: Assessment of children and adolescents*. New York: John Wiley and Sons.

Exner, J. E., Jr., & Weiner, I. B. (1994). *The Rorschach: A comprehensive system, volume 3: Assessment of children and adolescents*. New York: John Wiley and Sons.

Fiske, D. W., & Baughman, E. E. (1953). Relationships between Rorschach scoring categories and the total number of responses. *Journal of Abnormal & Social Psychology, 48*, 25-32.

Gacono, C. B., & Meloy, J. R. (1994). *The Rorschach assessment of aggressive and psychopathic personalities*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Goodman, S. N. (1999). Toward evidence-based medical statistics: I. The p value fallacy. *Annals of Internal Medicine, 130*, 995-1004.

Holtzman, W. H. (1958). *Holtzman inkblot technique*. San Antonio, TX, US: Psychological Corporation.

Jeffreys, H. (1961). *Theory of probability (3rd ed.)*. New York: Oxford University Press.

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association, 90*, 773-795.

Lindley, D. V. (1957). A statistical paradox. *Biometrika, 44*, 187–192.

Lipgar, R. M., & Waehler, C. A. (1991). A Rorschach investigation of mothers of behaviorally disturbed infants. *Journal of Personality Assessment, 56*, 106-117.

Mattlar, C. E. (2011). *The Issue of an Evolutionary Development of the Rorschach Comprehensive System (RCS) Versus a Revolutionary Change(R-PAS)*. http://www.rorschachtraining.com/category/articles/

Meyer, G. J. (1992). Response frequency problems in the Rorschach: Clinical and research implications with suggestions for the future. *Journal of Personality Assessment, 58(2)*, 231-244.

Meyer, G. J. (1993). The impact of response frequency on the Rorschach constellation indices and on their validity with diagnostic and MMPI-2 criteria. *Journal of Personality Assessment, 60*, 153–180.

Meyer, G. J., Erdberg, P., Shaffer, T. W. (2007). Toward international normative reference data for the Comprehensive System. *Journal of Personality Assessment, 89(S1)*, S201-S216.

Meyer, G. J., Viglione, D. J., Mihura, J. L., Erard, R. E., & Erdberg, P. (2011). *Rorschach Performance Assessment System: Administration, coding, interpretation and technical manual*. Toledo, OH: Rorschach Performance Assessment System.

Mihura, J. L., Meyer, G. J., Dumitrascu, N., & Bombel, G. (in press). The validity of individual Rorschach variables: Systematic reviews and meta-analyses of the Comprehensive System. *Psychological Bulletin*.

Perry, G. G., & Kinder, B. N. (1990). The susceptibility of the Rorschach to malingering: A critical review. *Journal of Personality Assessment, 54*, 47-57.

Rouder, J. N., & Morey, R. D. (2001). A Bayes factor meta-analysis of Bem's ESP claim. *Psychonomic Bulletin & Review, 18*, 682–289.

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t-tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review, 16*, 225–237.

Sellke, T., Bayarri, M. J., & Berger, J. O. (2001). Calibration of p values for testing precise null hypotheses. *American Statistician, 55*, 62-71.

Shaffer, T. W., Erdberg, P., & Haroian, J. (2007). Rorschach Comprehensive System data for a sample of 283 adult nonpatients from the United States. *Journal of Personality Assessment, 89 (S1)*, S166-S173.

Shrout, P. E., & Fliess, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin, 86*, 420-428.

Stanfill, M. L., Viglione, D. J., & Resende, A. C. (2013). Measuring psychological development with the Rorschach. *Journal of Personality Assessment, 95(2)*, 174-186.

Viglione, D. J. & Hilsenroth, M. J. (2001). The Rorschach: Facts, fictions, and future. *Psychological Assessment 13(4)*, 452-471.

Viglione, D. J., & Meyer G. J (2008). An overview of Rorschach psychometrics for forensic practice. In C. Gacono, F. Evans, N. Kaser-Boyd, & L. Gacono (Eds.) *The Handbook of Forensic Rorschach Assessment* (pp. 21-53). New York, NY: Routledge/Taylor & Francis Group.

Viglione, D. J., Blume-Marcovici, A., Miller, H., Giromini, L., & Meyer G. J. (2012). An initial inter-rater reliability study for the Rorschach performance assessment system. *Journal of Personality Assessment*, *94 (6),* 607–612, DOI:10.1080/00223891.2012.684118.

Viglione, D. J., Giromini, L., Gustafson, M., & Meyer, G. J. (2012). Developing continuous variable composites for Rorschach measures of thought problems, vigilance, and suicide risk. *Assessment*. DOI: 10.1177/1073191112446963.

Viglione, D. J., Perry, W., Giromini, L., & Meyer, G. J. (2011). Revising the Rorschach Ego Impairment Index to accommodate recent recommendations about improving Rorschach validity. *Journal of Personality Assessment, 11(4)*, 349-364.

Wagenmakers, E.-J. (2007). A practical solution to the pervasive problem of p values. *Psychonomic Bulletin &Review, 14*, 779–804.

Wagenmakers, E.-J., & Grünwald, P. (2006). A Bayesian perspective on hypothesis testing: A comment on Killeen (2005). *Psychological Science, 17*, 641-642.

Wechsler, D. (2008). *Wechsler Adult Intelligence Scale-Fourth Edition: Technical and interpretive manual*. San Antonio, TX: Pearson.

Weiner, I. B. (2003). *Principles of Rorschach interpretation*. New York: Routledge/Taylor & Francis Group.

Weiner, I. (2012). *Rorschach Training Programs, INC Newsletter, Volume 4; Number 5*. http://www.rorschachtraining.com/category/newsletters/

Table 1. Composition of the Samples.

	CS (N = 142)	R-PAS (N = 99)
Gender (Phi = .06; p = .33)		
Boys	86 (60.6%)	66 (66.7%)
Girls	56 (39.4%)	33 (33.3%)
Age (t (239) = -.80; p = .42)		
Mean	12.2	12.5
SD	3.0	3.0
Years of Education (t (239) = -.89; p = .37)		
Mean	6.7	7.0
SD	3.0	2.9
Ethnicity (χ^2 = .88; p =.65)		
Caucasian	112 (78.9%)	74 (74.7%)
African-American	17 (12.0%)	16 (16.2%)
Other	13 (9.2%)	9 (9.1%)
SES (Mann-Whitney U = 6946.5; p = .87)		
Lower-Lower	63 (44.4%)	47 (47.5%)
Lower-Middle	1 (0.7%)	0 (0%)
Lower-Upper	24 (16.9%)	13 (13.1%)
Middle-Lower	26 (18.3%)	18 (18.2%)
Middle-Middle	20 (14.1%)	17 (17.2%)
Middle-Upper	8 (5.6%)	4 (4.0%)
Diagnostic Category (χ^2 = 17.67; p =.01) [(a)]		
ADHD	26 (18.3%)	19 (19.2%)
Anxiety Disorder (excluding PTSD)	12 (8.5%)	16 (16.2%)
Autism Spectrum Disorder	7 (4.9%)	4 (4.0%)
Behavior Disorder (excluding ADHD)	12 (8.5%)	5 (5.1%)
Mood Disorder	42 (29.6%)	20 (20.2%)
Psychotic Disorder	2 (1.4%)	0 (0.0%)
PTSD	26 (18.3%)	10 (10.1%)
Sexual Behavior (Court-Involved)	15 (10.6%)	25 (25.3%)

Notes. (a) Although this test was significant, examination of the standardized residuals indicated that the two groups only differed as for the presence of court-involved sexual behavior problem children. When excluding this subgroup from the analysis, in fact, no other diagnostic differences approached statistical significance. Follow up analyses indicated that this difference did not account for the main results of the study.

Table 2. Inter-rater Reliability of the 51 Selected Rorschach Variables.

R-PAS Page 1 Variables	ICC	R-PAS Page 2 Variables	ICC
Engagement & Cognitive Processing		*Engagement & Cognitive Processing*	
Complexity	0.99	W%	1.00
R	1.00	Dd%	1.00
F%	0.96	IntCont	0.85
Blend	0.95	Vg%	0.72
Sy	0.89	V	0.96
MC	0.90	FD	0.89
MC – PPD	0.96	R8910%	1.00
M	0.88	WsumC	0.95
M/MC	0.91	C	0.90
(CF+C)/SumC	0.86	Mp/(Ma+Mp)	0.66
Perception & Thinking Problems		*Perception & Thinking Problems*	
EII-3	0.94	FQu%	0.72
TP-Comp	0.94		
WsumCog	0.96		
SevCog	0.90		
FQ-%	0.89		
WD-%	0.88		
FQo%	0.92		
P	0.84		
Stress & Distress		*Stress & Distress*	
m	0.94	PPD	0.99
Y	0.91	YTVC'	1.00
MOR	0.85	Cblend	0.89
SC-Comp	0.94	C’	0.97
		CritCont%	0.98
Self & Other Representation		*Self & Other Representation*	
PHR/GPHR	0.73	SumH	0.95
M-	0.71	NPH/SumH	0.87
V-Comp	0.91	r	1.00
H	0.99	p/(a+p)	0.60
COP	0.86	AGM	0.95
		T	0.70
		PER	0.97
		An	1.00

Table 3. Distribution of R within CS and R-PAS Administrations.

Administration		CS (n = 142)	R-PAS (n = 99)
Mean		24.3	24.7
Mode		20	20
S.D.		8.4	5.4
Min.		14	14
Max.		53	40
Percentiles:	5	14.0	18.0
	10	16.0	20.0
	50	22.0	23.0
	90	35.7	32.0
	95	41.0	38.0
% of Low-R records (R < 17)		15.5 %	1.0 %
% of High-R records (R > 27)		29.6 %	25.3 %

Table 4. Mean Values of Page 1 R-PAS Variables Produced by CS vs. R-PAS Administrations.

R-PAS Page 1 Variables	CS (N = 142) M	CS (N = 142) SD	R-PAS (N = 99) M	R-PAS (N = 99) SD	$t_{(239)}$	p	d	JZS B
Engagement & Cognitive Processing								
Complexity	69.1	28.0	67.9	21.6	.3	.74	.04	4.9
R	24.3	8.4	24.7	5.4	-.5[(b)]	.62	-.06	4.5
F%	53.5	17.1	53.8	18.5	-.1	.88	-.02	5.1
Blend	3.1	2.8	3.1	2.4	.2	.86	.02	5.0
Sy	6.0	4.0	5.4	3.8	1.1	.26	.15	2.9
MC	5.4	3.6	4.8	3.1	1.3	.20	.17	2.4
MC – PPD	-2.3	4.6	-3.4	4.6	1.7	.08	.23	1.4
M	2.5	2.4	2.3	2.2	.5	.60	.07	4.5
M/MC	42.7	26.0	46.7	28.2	-1.0[(b)]	.30	-.15	3.0
(CF+C)/SumC	49.9	28.3	39.8	31.6	2.0[(b)]	.05	.34	0.7
Perception & Thinking Problems								
EII-3[(a)]	.7	1.2	.7	1.4	-.2	.83	-.03	5.0
TP-Comp[(a)]	1.5	1.3	1.5	1.5	.2	.86	.02	5.0
WsumCog	15.5	16.7	18.9	20.4	-1.1[(c)]	.26	-.18	2.9
SevCog	.8	1.5	1.1	2.1	-.4[(c)]	.67	-.15	4.7
FQ-%[(a)]	27.7	12.5	27.2	13.6	.3	.73	.04	4.9
WD-%[(a)]	24.2	13.9	23.5	13.5	.4	.69	.05	4.7
FQo%[(a)]	39.3	12.9	39.9	13.9	-.4	.72	-.05	4.7
P	3.2	1.5	3.3	1.5	-.8	.42	-.11	3.8
Stress & Distress								

m	1.4	1.8	1.7	1.8	-1.4	.17	-.18	2.1
Y	1.3	1.8	1.1	1.5	.8	.40	.11	3.8
MOR	2.3	2.2	2.3	2.3	.0	.97	.01	5.1
SC-Comp[(a)]	5.2	1.3	5.1	1.3	.9	.38	.11	3.5
Self & Other Representation								
PHR/GPHR	56.9	23.0	53.8	23.8	1.0[(b)]	.33	.14	3.1
M-[(a)]	.7	1.2	.6	1.0	.5[(c)]	.87	.06	4.5
V-Comp[(a)]	3.7	1.5	3.4	1.3	1.3	.21	.16	2.4
H	1.9	1.8	2.1	1.7	-1.2	.24	-.15	2.7
COP	.4	.7	.3	.6	1.1[(b)]	.29	.13	2.9

Notes. (a) The index was generated using the CS FQ tables, which are slightly different from the R-PAS FQ tables; (b) Degrees of freedom were lower than 239 because either some missing values were present or homoscedasticity could not be assumed and Welch-Satterthwaite method was used; (c) The t-test was computed after mathematically transforming the variable because of nonnormality issues.

Table 5. Mean Values of Page 2 R-PAS Variables Produced by CS vs. R-PAS Administrations.

R-PAS Page 2 Variables	CS (N = 142)		R-PAS (N = 99)		$t_{(239)}$	p	d	JZS B
	M	SD	M	SD				
Engagement & Cognitive Processing								
W%	36.8	22.0	35.0	20.7	.6	.52	.08	1.4
Dd%	29.2	15.4	28.9	14.9	.1	.89	.02	5.1
IntCont	1.3	1.8	1.4	1.9	-.3[c]	.73	-.05	4.9
Vg%	1.7	3.9	1.8	4.1	-.2[c]	.87	-.02	5.0
V	.3	1.0	.3	1.0	.4[c]	.72	.01	4.7
FD	.7	1.1	.9	1.2	-1.0	.31	-.13	3.2
R8910%	32.1	8.8	31.8	5.3	.4[b]	.72	.04	4.7
WsumC	2.9	2.1	2.5	2.0	1.5	.13	.20	1.8
C	.4	.7	.3	.6	1.1[c]	.29	.15	2.9
Mp/(Ma+Mp)	58.9	29.6	49.5	24.7	1.7[b]	.10	.34	1.1
Perception & Thinking Problems								
FQu%[a]	31.3	11.1	31.1	11.4	.2	.86	.02	5.0
Stress & Distress								
PPD	7.7	5.0	8.2	5.1	-.7	.47	-.09	4.1
YTVC'	4.1	3.4	4.2	3.8	-.1	.89	-.02	5.1
Cblend	.7	.9	.6	.8	.7	.48	.09	4.1
C’	2.3	2.1	2.5	2.7	.0[c]	>.99	-.06	5.1
CritCont%	25.6	20.1	23.1	17.4	1.2	.25	.15	2.7
Self & Other Representation								
SumH	5.7	3.2	5.7	3.0	.0	>.99	.00	5.1
NPH/SumH	68.1	24.1	61.5	26.2	1.9[b]	.06	.26	1.0

r	.3	.7	.2	.7	1.3[(c)]	.21	.08	2.4
p/(a+p)	50.8	23.6	46.7	21.7	1.2[(b)]	.22	.18	2.7
AGM	.5	1.0	.5	.8	-.4[(c)]	.73	.00	4.7
T	.2	.5	.2	.4	.3[(c)]	.78	.06	4.9
PER	1.7	2.2	1.1	1.7	2.1[(b)]	.04	.26	.7
An	1.4	1.7	1.2	1.8	1.2[(c)]	.25	.14	2.7

Notes. (a) The index was generated using the CS FQ tables, which are slightly different from the R-PAS FQ tables; (b) Degrees of freedom were lower than 239 because either some missing values were present or homoscedasticity could not be assumed and Welch-Satterthwaite method was used; (c) The t-test was computed after mathematically transforming the variable because of nonnormality issues.

Table 6. Impact of Administration Method on Complexity: Mean Values of Rorschach Variables Related to Complexity when Including and Excluding the 1st Responses to Each Card.

Complexity Variables	CS (N = 142)		R-PAS (N = 99)		$t_{(239)}$	p	d	JZS B
	M	SD	M	SD				
Including the 1st Responses								
R	24.3	8.4	24.7	5.4	-.5[a]	.62	-.06	4.5
Location-Developmental Quality Complexity	33.1	11.9	32.3	8.6	.6[a]	.53	.08	4.3
Determinant Complexity	15.0	8.3	15.0	7.6	-.1	.94	-.01	5.1
Content Complexity	21.0	9.9	20.6	8.1	.3	.76	.04	4.9
Response Complexity (Complexity/R)	2.9	.8	2.8	.7	1.4	.17	.18	2.1
Pure F	13.2	6.5	13.2	5.2	-.1[a]	.95	-.01	5.1
Common D Location	8.7	5.8	9.1	4.8	-.7	.51	-.09	4.1
Animal Content (A)	8.5	4.1	9.1	3.9	-1.1	.26	-.15	2.9
P	3.2	1.5	3.3	1.5	-.8	.42	-.07	3.8
Sy	6.0	4.0	5.4	3.8	1.1	.26	.15	2.9
Blend	3.1	2.8	3.1	2.4	.2	.86	.00	5.0

Excluding the 1st Responses								
R	14.3	8.4	14.7	5.4	-.5(a)	.62	-.06	4.5
Location-Developmental Quality Complexity	18.5	11.4	18.2	7.6	.2(a)	.86	.02	5.0
Determinant Complexity	8.0	6.3	8.7	5.7	-.9	.35	-.12	3.5
Content Complexity	11.7	8.5	11.8	6.1	.0(a)	.96	-.01	5.1
Response Complexity (Complexity/R)	2.7	.8	2.6	.7	.4	.66	.06	4.7
Pure F	8.1	5.6	7.9	4.2	.3(a)	.77	.04	4.9
Common D Location	5.8	4.7	6.1	3.8	-.5(a)	.63	-.06	4.5
Animal Content (A)	4.7	3.6	4.8	2.9	-.4	.72	-.05	4.7
P	.9	1.0	.9	1.0	-.7	.51	.00	4.1
Sy	3.0	3.0	2.7	2.6	-.8	.43	.11	3.8
Blend	1.5	1.8	1.7	1.8	-1.0	.32	.11	3.2

Notes. (a) Degrees of freedom were lower than 239 because homoscedasticity could not be assumed and Welch-Satterthwaite method was used.